1303113

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

March 31, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/489,317*
FILING DATE: *July 22, 2003*
RELATED PCT APPLICATION NUMBER: *PCT/US04/23566*

Certified by

Jon W. Dudas

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

Mail Stop Provisional Patent Application
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Transmitted herewith for filing is a pr visional patent application under CFR 1.53(c) of Inventors:

First Name and Middle Name [if any] followed by Last Name	RESIDENCE (City/State/Country)
Imran Shah	**Denver/Colorado/USA**
Daniel Charles McShan	**Littleton/Colorado/USA**

Title: METHODS, SYSTEMS, AND SOFTWARE FOR PREDICTING BIOCHEMICAL PATHWAYS

Enclosed are:

[X] 24 pages of the application (including description, claims, abstract and table).
[X] 08 sheet(s) of [ ] formal [x] informal drawing(s).
[X] Abstract.
[X] 03 claims.
[X] Small Entity Status is Claimed
[X] Cover Sheet.
[X] Patent Application Filing Acknowledgement postcard.
[X] the invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.
[X] Application Data Entry Sheet.

Please charge Deposit Account No. 50-0893 as follows:

[X] Filing fee. $ $80.00
[X] Any additional fees associated with this paper or during the pendency of this application.

1 extra copy of this sheet is enclosed.

Respectfully submitted,

Christopher C.Sappenfield
Reg No.: 45,075
Applicant's Representative

CORRESPONDENCE ADDRESS:
QUINE INTELLECTUAL PROPERTY LAW GROUP, P.C.
P.O. Box 458
Alameda, CA 94501
Telephone: (510) 337-7871
Fax: (510) 337-7877

Customer No. 22798
22798
PATENT TRADEMARK OFFICE

Attorney Docket No. 61-000800US
Client Reference No. CU1113H
Express Mail Label No. EV 297621969 US
Date of Deposit: July 22, 2003

I hereby certify that this is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above, addressed to: Mail Stop Provisional Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450

By: Chanti Appling

"Express Mail" Label No. EV 297621969 US

Date of Deposit: July 22, 2003

I hereby certify that this is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above, addressed to: Mail Stop Provisional Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450

By: Chranti Appling

Docket No.: 61-000800US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

PROVISIONAL U.S. PATENT APPLICATION FOR

METHODS, SYSTEMS, AND SOFTWARE FOR PREDICTING BIOCHEMICAL PATHWAYS

Inventors: **IMRAN SHAH**, a citizen of the United States of America, residing at 1230 Ivanhoe Street, Denver, Colorado 80220.

DANIEL CHARLES MCSHAN, a citizen of the United States of America, residing at 2901-D West Long Drive, Littleton, Colorado 80120.

Assignee: **THE REGENTS OF THE UNIVERSITY OF COLORADO**, 201 Regent Administrative Center, 3 SYS, Boulder, Colorado 80309

Entity: **SMALL**

As Filed: **TUESDAY, JULY 22, 2003**

Correspondence Address:

QUINE INTELLECTUAL PROPERTY LAW GROUP, P.C.

P.O. Box 458
Alameda, CA 94501
Internet address: www.quinelaw.com

Phone: (510) 337-7871
Fax: (510) 337-7877
Email: jaquine@quinelaw.com

METHODS, SYSTEMS, AND SOFTWARE FOR PREDICTING BIOCHEMICAL PATHWAYS

COPYRIGHT NOTIFICATION

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

[0002] This invention was made with government support under Grant No. BES-9911447 awarded by the National Science Foundation, Grant No. DE-FG03-01ER63111 awarded by the Department of Energy, and Grant No. N00014-00-1-0749 awarded by the Office of Naval Research. The government may have certain rights in the invention.

FIELD OF THE INVENTION

[0003] The present invention relates generally to predicting or inferring biochemical pathways. In particular, the invention provides methods, systems, and computer program products for automatic biochemical pathway inference.

BACKGROUND OF THE INVENTION

[0004] Automated methods for biochemical pathway inference or prediction are becoming increasingly important for understanding biological processes in living and synthetic systems. With the availability of data on complete genomes and increasing information about enzyme-catalyzed biochemistry it is becoming feasible to approach this problem computationally. However, even with the availability of a genomic blueprint for a living system and functional annotations for its putative genes, the experimental elucidation of its biochemical processes is typically still a daunting task.

Though it is possible to organize genes by broad functional roles, piecing them together manually into consistent biochemical pathways can quickly become intractable.

[0005] A number of metabolic pathway reconstruction tools have been alleged since the availability of the first microbial genome, *H. influenza* (Fleischmann et al. (1995) "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd," Science 269:469-512). These include PathoLogic (Karp & Riley, Representations of metabolic knowledge: Pathways. In Second International Conference on Intelligent Systems for Molecular Biology (Altman, R., Brutlag, D., Karp, P., Lathrop, R. & Searls, D., eds), AAAI Press (1994)), MAGPIE (Gaasterland & Selkov (1995) "Automatic Reconstruction of Metabolic Networks Using Incomplete Information," Intelligent Systems for Molecular Biology 3:127-135 and Gaasterland & Sensen (1996) "MAGPIE: automated genome interpretation," Trends Genet 12(2):76-78), WIT (Overbeek et al. (2000) "Wit: integrated system for high-throughput genome sequence analysis and metabolic reconstruction," Nucleic Acids Res 28(1):123-125) and PathFinder (Goesmarm et al. (2002) "PathFinder: reconstruction and dynamic visualization of metabolic pathways," Bioinformatics 18(1):124-129). The goal of most pathway inference methods has generally been to match putatively identified enzymes with known, or "reference", pathways. Although reconstruction can be a useful starting point for elucidating the metabolic capabilities of an organism based upon prior pathway knowledge, reconstructed pathways often have many missing enzymes, even in essential pathways.

[0006] In addition, the issue of redefining microbial biochemical pathways based on "missing" enzymes is often of consequence since there are many examples of alternatives to standard pathways in a variety of organisms (Cordwell (1999) "Microbial genomes and missing enzymes: redefining biochemical pathways," Arch Microbiol 172(5):269-279). Moreover, engineering a new pathway into an organism through, e.g., heterologous enzymes also requires the ability to infer new biochemical routes.

SUMMARY OF THE INVENTION

[0007] In one aspect, the invention relates to a method of predicting a biochemical pathway (e.g., a metabolic pathway, such as an anabolic or catabolic pathway). The method includes providing a population of compounds. The population comprises one or more input compounds and one or more output compounds. The

method also includes defining at least one state-space that comprises the population of compounds. In addition, the method also includes identifying one or more candidate biochemical pathways between at least one of the input compounds and at least one of the output compounds using at least one informed search technique (e.g., a heuristic search technique, etc.) to search the state-space.

[0008] In another aspect, the invention provides a computer program product comprising a computer readable medium having one or more logic instructions for receiving data that defines at least one state-space comprising a population of compounds, which population comprises one or more input compounds and one or more output compounds. The computer readable medium also include one or more logic instructions for identifying one or more candidate biochemical pathways between at least one of the input compounds and at least one of the output compounds using at least one informed search technique to search the state-space.

[0009] In still another aspect, the invention relates to a system for predicting a biochemical pathway. The system comprises at least one computer having system software comprising one or more logic instructions for receiving data that defines at least one state-space comprising a population of compounds, which population comprises one or more input compounds and one or more output compounds. The system software also comprises one or more logic instructions for identifying one or more candidate biochemical pathways between at least one of the input compounds and at least one of the output compounds using at least one informed search technique to search the state-space.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] Figure 1 is a flow chart illustrating a method of predicting biochemical pathways according to specific embodiments of the invention.

[0011] Figure 2 provides an alphabetical list of 145 descriptors used to represent chemical state-space according to one embodiment of the invention. As shown, atoms are represented by their IUPAC symbols. Single, double, and triple bonds are represented as the symbols -, =, and #, respectively.

[0012] Figure 3 schematically depicts certain known chemical successors of α-D-glucose (adg), denoted T^{adg}. More specifically, the schematically illustrated

structure of adg is shown on the left side of the figure, while the structures of the successors are schematically shown on the right side of the figure.

[0013] Figure 4 schematically shows a best-first search algorithm to find pathway, $P^{0,L}$, from an input compound, x^0, to output compound x^L, using a heuristic evaluation function, F.

[0014] Figure 5 A and B illustrate example interfaces for predicting biochemical pathways using a computer interface, possibly over a web page, according to specific embodiments of the present invention.

[0015] Figure 6 is a block diagram showing a representative example logic device in which various aspects of the present invention may be embodied.

[0016] Figure 7 is a block diagram illustrating an integrated system according to specific embodiments of the present invention.

[0017] Figure 8 schematically shows the visualization of a linear pathway.

DETAILED DISCUSSION OF THE INVENTION

I. DEFINITIONS

[0018] Before describing the present invention in detail, it is to be understood that this invention is not limited to particular methods, systems, computers, or computer readable media, which can, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting. Further, unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. In describing and claiming the present invention, the following terminology and grammatical variants will be used in accordance with the definitions set forth below.

[0019] A "population" refers to a collection of at least two molecule or compound types, e.g., 2, 3, 4, 5, 10, 20, 50, 100, 1,000 or more molecule or compound types.

[0020] A "descriptor" refers to something that serves to describe or identify an item. For example, chemical descriptors can be used to describe a compound in terms of, e.g., the number and/or types of constituent atoms of the compound, the number and/or types bonds of the compound, and/or other attributes of the compound.

[0021] A "biochemical pathway" refers to a biochemical reaction or sequence of biochemical reactions that begins with one or more input compounds and yields one or more output compounds. One or more reaction steps in a biochemical pathway are typically catalyzed by one or more biocatalysts.

[0022] A "biocatalyst" refers to a catalyst that reduces the activation energy of a biochemical reaction involving input and output compounds. Exemplary biocatalysts include enzymes, which are protein- and/or nucleic acid-based catalysts.

[0023] An "input compound" or "initial compound" refers to a reactant or a representation of a reactant in a given chemical reaction. An "output compound," "destination compound," or "successor compound" refers to a product or a representation of a product in a given chemical reaction.

[0024] The term "state-space" refers a population of states (e.g., chemical compounds, etc.) and to transitions (e.g., chemical reactions, etc.) between those states in the population.

II. BIOCHEMICAL PATHWAY PREDICTION

[0025] The methods of the invention predict biochemical routes by reasoning over transformations using chemical and biological information. More specifically, the present invention provides computational approaches for automated pathway prediction that are useful for exploring plausible biochemical routes underlying various biological processes. Although essentially any programming language is optionally utilized to implement the methods described herein, certain specific embodiments referred to herein are implemented in Common Lisp (*see, e.g.,* Graham, ANSI Common LISP, 1st Ed., Prentice Hall (1995) and Norvig, Paradigms of Artificial Intelligence Programming: Case Studies in Common Lisp, Morgan Kaufmann (1991). In addition, a flexible web-based interactive system, called PathMiner, that embodies aspects of the present invention is described further in, e.g., McShan et al. (2003) "PathMiner: Predicting Metabolic Pathways by Heuristic Search," Bioinformatics (in press). There are at least two broad biological applications of the invention. First, to investigate pathways in an organism using information about its functionally characterized proteins or other biocatalysts. Second, to synthesize novel pathways for engineering new biochemical capabilities.

[0026] Going beyond standard pathways is one objective of the present invention, which uses chemically motivated heuristics to guide the search for pathways. As such, it complements pre-existing approaches like PathoLogic (referred to above) which find the best candidate reference pathways and the corresponding genes in a living system. Other approaches to pathway synthesis have also included the work of Sesseriotis and Bailey (Seressiotis & Bailey (1988) "Mps: an artificially intelligent software system for the analysis and synthesis of metabolic pathways," Biotechnology and Bioengineering 31:587-602), and later, Mavrovouniotis's approach for pathway generation based on the consideration of thermodynamic feasibility of reactions (Mavrovouniotis, "Identification of Qualitatively Feasible Metabolic Pathways" In: Artificial Intelligence and Molecular Biology, (Hunter (Ed.)) AAAI (1993)). Further, the present invention can be used interactively to search for biochemical routes in the context of specific organisms or to identify synthetic pathways.

[0027] In overview, the present invention abstracts biochemical processes in terms of a biochemical state-space: compounds define the states and transformations between compounds define the state-transitions. Pathway prediction is then considered as a problem of searching the biochemical state-space. State-space and an embodiment of an algorithm for predicting pathways through search are described below. To further illustrate, Figure 1 provides a flow chart illustrating an example method according to specific embodiments of the invention. As shown, the method includes as follows: providing a population of compounds having input and output compounds (A1); defining a state-space that comprises the population of compounds (A2); and identifying a candidate biochemical pathway between an input compound and an output compound using a heuristic search technique to search the state-space (A3).

A. THE BIOCHEMICAL STATE-SPACE

[0028] The notion of the biochemical state-space is based on resolving enzyme-catalyzed biochemistry into two components. The first component is the chemical component, which represents transformations between, e.g., metabolites. The second component is the biocatalytic component, which involves transformations catalyzed by enzymes or other biocatalysts. This logical separation between biocatalysts and the chemistry they catalyze is evolutionarily plausible and functionally relevant,

since a biocatalyst can often catalyze multiple transformations. Additional details relating to abstracting the interaction of a biocatalyst with a chemical transformation are provided in, e.g., Karp (Karp & Riley (1994), *supra*). By considering chemical transformations and biocatalysts separately, they can be dealt with rationally to infer plausible pathways.

1. COMPOUNDS

[0029] The present invention includes defining a simple representation for compounds that captures their essential chemical properties, which are available from, e.g., existing sources of data. In certain embodiments, for example, a compound is denoted as x in state-space and described by a set of chemical descriptors, x_k. Thus, every compound can be placed at a point in hyperspace, which is defined by $x = (x_1, x_2, x_3,..., x_N)$. For example, compounds can be described based on the composition of their atoms and bonds. To illustrate, the embodiment depicted in Figure 2 includes a total of 145 unique features. More specifically, based on the 145 descriptors in Figure 2, α-D-glucose (adg) is represented as the vector $x^{adg} = (0, 0, 0, 0, 0, 6, 0, 0, 0, ...)$. Similarly, pyruvate (pyr) is represented as the vector $x^{pyr} = (0, 0, 0, 0, 0, 3, 0, 0, 0, ...)$. Since the chemical descriptor space is large and the vector for any given compound is sparse, compound vectors can typically be succinctly expressed using an attribute-value notation. In this notation carbon dioxide, x^{CO2}, is described in equation 1.

$$x^{CO2} = ((C\ 1)(O\ 2)(C = O\ 2)) \qquad (1)$$

Equation 1 states that CO_2 is defined by the state vector, x^{CO2}, which contains three components: the number of carbon atoms ($x_C = 1$), the number of oxygen atoms ($x_O = 2$), and the number of C = O bonds ($x_{C=O} = 2$). The set of 145 descriptors represents most compounds uniquely. However, since chirality is not represented in this embodiment, stereoisomers map to identical points in the state-space. For example, the representation of β-D-glucose and α-D-glucose are identical. In other embodiments of the invention, descriptors representing chirality are included to account for stereoisomers.

2. TRANSFORMATIONS

[0030] To represent transformations, the complex bond changes that occur when one compound is converted to another are approximated. Transformations are abstracted as transitions between compound states denoted by t. For example, the transformation of α-D-glucose (x^{adg}) into α-D-glucose-6-phosphate (x^{adg6P}) (i.e., $x^{adg} \rightarrow x^{adg6P}$) can be defined as the vector difference, which is shown in equation 2.

$$\begin{aligned} t^{adg,adg6P} &= x^{adg6P} - x^{adg} \\ &= ((C6)(H12)(O10)(P1)\ldots) - \\ &\quad ((C6)(H12)(O6)(P0)\ldots) \\ &= ((P1)(O4)(P-O3)) \end{aligned} \qquad (2)$$

In equation 2, the term $t^{adg,adg6P}$ describes the state-transition as the addition of ((*P*1)(*O*4)(*P* - *O*3)). This set of descriptors corresponds to a known chemical moiety, which is the phosphate functional group (PO_4^{3-}).

[0031] Though it is not be possible to do so in all cases, the interpretation of state-transitions is often chemically intuitive. Each compound can be chemically transformed into a number of other "successor" compounds. For example, some of the chemical successors of α -D-glucose are shown in Figure 3. As described further below, the set of known chemical successors of x can be denoted as T.

[0032] Each state-transition can typically be related to a known biocatalyst. In some embodiments, for example, biochemical transformations are considered as state-transitions, t, that approximate the complex bond changes in, e.g., enzyme-catalyzed reactions. Optionally, additional biochemical attributes of reactions, like structural and energetic changes, are also included in these representations.

B. PATHWAY PREDICTION AS SEARCH

[0033] After defining compounds as states and transformations as state-transitions, pathway inference or prediction becomes a state-space search problem. State-space searches have been referred to in, e.g., Artificial Intelligence (AI) research (*see, e.g.,* Pearl, Heuristics: Intelligent Search Strategies for Computer Problem Solving AddisonWesley (1984)). The present invention addresses the problem of predicting a biochemical pathway as searching a route from an initial compound to a destination compound through a series of state-transitions. In particular, the initial or input

compound is denoted as x^0, the destination or output compound is denoted as x^L, and the pathway between these two is denoted as $P^{0,L}$. This is further shown in equation 3.

$$P^{0,L} = x^0 \rightarrow x^1 \rightarrow x^2 \rightarrow \ldots x^m \ldots \rightarrow x^L \qquad (3)$$

[0034] The simplest approach for pathway inference is an uninformed search, which includes depth-first search and breadth-first search methods. In uninformed searches, successive states from x^0 are explored blindly until the goal, x^L, is reached. In real-world problems, like biochemical pathway searches, blind searches can lead to a combinatorially large number of possible solutions. For practical purposes it is typically desirable to reduce the set of solutions to a smaller subset. To address this issue, informed search techniques can reason over the state-space to infer pathways that satisfy some optimality condition. For example, a heuristic search is an informed search technique that can systematically explore a state-space by measuring the cost associated with any state-transition (Pearl (1984), *supra*). Informed searches generally take the form of best-first searches that use a heuristic evaluation function, called *F*, to reduce the combinatorially large number of possibilities faced by other methods. A simplified version of a best-first search is given in the algorithm provided in Figure 4.

[0035] More specifically, Figure 4 schematically shows a best-first search algorithm to find pathway, $P^{0,L}$, from an input compound, x^0, to output compound x^L, using a heuristic evaluation function, *F*. As shown, in each iteration the successors, T, of the best state in the list *X* are explored as follows. If the goal is reached then the search terminates with a path, $P^{0,L}$. Otherwise, there are two options. First, if the state x^m is not in *X*, then it is added to it using push(x^m,*X*), and a pointer from the state to its parent is created with point(x^m,x). Second, if the state is in *X* then its heuristic score is updated to the lower out of the current and old values. Each state in *X* points to its predecessors and the path from x^0 can be traced using path(x^m). The search terminates when there are no more states to explore.

[0036] The heuristic evaluation function, *F*, can be calculated using different methods. For example, a greedy search minimizes the cost of reaching the goal state from the current state (called *H*). Conversely, a uniform cost search minimizes the cost of reaching the current state from the initial state (called *G*). Certain embodiments of the invention utilize A* (A-star) searches, which use an evaluation function that is the

sum of the estimated cost thus far (G) and the estimated cost to the goal (H). In effect, this minimizes the overall path cost ($F = H + G$). To infer biochemical pathways by heuristic search, the present invention includes a strategy for calculating the cost of a pathway, as described below.

1. HEUREKA

[0037] This section relates to defining the cost of state changes in a state-space representation as described herein. The biological factors that determine the cost of a pathway in a living system are not always known. Evolution, environment, bioenergetics, kinetics, growth, or a broader biochemical context, all may contribute to the existence of a biochemical pathway in an organism. The problem is that it can be difficult to calculate the contribution of most of these factors due to the scarcity of data, or the limitations of current knowledge. In certain embodiments of the invention, state-space is used to define the cost based on the chemical efficiency of a pathway. While this may not always be biologically correct, it is congruent with the notion that living systems tend to optimize their growth. Furthermore, it is a useful heuristic for finding synthetic pathways.

[0038] To formalize the notion of cost in state-space, the difference between any two compounds is defined as Δx and the corresponding distance as $|\Delta x|$. For a state-transition this is simply $t = \Delta x$. The distance, $|\Delta x|$, can be calculated using, e.g., the Manhattan metric or the Euclidean metric, which are given in equations 5 and 4, respectively. Either the Manhattan distance or the Euclidean distance is admissible as a heuristic, because it represents the shortest distance between any two compounds. The Manhattan distance (equation 5) is often used, because the discrete chemical changes are typically more intuitive and the computation is typically more efficient than with other distances.

$$|\Delta \mathbf{x}|_E = \sqrt{\sum_{k=0}^{k=N} (\Delta x_k)^2} \qquad (4)$$

$$|\Delta \mathbf{x}|_M = \sum_{k=0}^{k=N} \Delta x_k \qquad (5)$$

[0039] Using the notion of distance between states, the functions *F*, *G* and *H*, which are utilized for heuristic searches, can be evaluated. To illustrate with the hypothetical pathway shown in equation 3 (described above), which begins with the initial state, x^0, ends with the final state, x^L, and has any intermediate state, x^m. The calculation of the cost functions *G* and *H* at the intermediate state, x^m, is given in equations 6 and 7, respectively.

$$G(0,m) = \sum_{i=1}^{i=m} |x^i - x^{i-1}| \qquad (6)$$

$$H(m,L) = |x^m - x^L| \qquad (7)$$

[0040] For an A* search the state selected for further exploration minimizes the total cost, $F = G + H$, which is shown in equation 8. Intuitively, $G(0,m)$ is the actual distance due to chemical transitions from x^0 to x^m, whereas $H(m, L)$ *is* a "guess" for the shortest possible distance to the goal state x^L.

$$\begin{aligned} F(0,m,L) &= G(0,m) + H(m,L) \qquad (8) \\ &= \sum_{i=1}^{i=m} (|x^i - x^{i-1}|) + |x^m - x^L| \end{aligned}$$

[0041] The intuition behind a heuristic search for a pathway includes, for example, as follows. One wants to find the series of efficient biochemical transformations that convert one compound into another. In state-space, the heuristic (*H*) is a guide for the chemical proximity of any intermediate state to the goal. By using the evaluation function in equation 8 in the algorithm schematically shown in Figure 4, one can select the pathway that efficiently converts the input to the output. The efficiency of this conversion is typically not determined by the length of the pathway. Rather, it is generally defined by an optimal value for the heuristic evaluation function, *F*. Although in certain embodiments of the invention this function is calculated in terms of chemical distance, any biochemical property that can be calculated from available biochemical information is also optionally utilized.

[0042] The search can also be guided by, e.g., using biological information. To illustrate, pathways can be searched for in an organism using a list of enzymes annotated in the genomic sequence. This can be accomplished, e.g., by modifying the algorithm schematically shown in Figure 4 to alter the successors for each

state (at statement, T ← successors(x)). Since each state-transition is associated with an enzyme, the available state-transitions and allowed successors for each state are constrained by the available enzymes.

2. EVALUATING EFFICIENCY

[0043] In order to evaluate the efficiency of different search methods for computing each pathway one can calculate the effective branching factor, called b^*. The branching factor, called b, is the number of successors for a given state. The effective branching factor for a given computed pathway of length L with M nodes expanded is defined as the branching factor that a uniform tree of depth d would possess in order to contain M states. The relationship between M,d,b^* can be expressed by the polynomial given in equation 9, which can be solved numerically to estimate b^*.

$$M = \sum_{j=0}^{j=L} (b*)^j \qquad (9)$$

3. DATA

[0044] The methods, systems, and software described herein can use compound, transformation, and enzyme information from essentially any source. In one embodiment of the invention, for example, the Kyoto Encyclopedia of Genes and Genomes (KEGG) is used at least in part due to its accessibility and breadth of biochemical data. Parsers have been developed to import KEGG data into Lisp (the programming language in which certain embodiments of the invention are implemented). To populate biochemical state-space compound data is optionally extracted from KEGG. In certain embodiments, state-transitions are further refined to use only, e.g., the main substrates in transformations (or main transformations). To illustrate with the following reaction:

$$Ethanol + NAD^+ \leftrightharpoons Acetaldehyde + NADH + H^+$$

For example, one can map this reaction to the state-transition $Ethanol \leftrightharpoons Acetaldehyde$, but not to $Ethanol \leftrightharpoons H^+$. In some embodiments, algorithmic approaches to decomposing reactions into substrate and product relations can be utilized. In others embodiments, KEGG pathway maps or the like which already contain "main" reactions are optionally utilized. For example, this data is available from the KEGG pathway map files, which

contain unordered lists of the main transformations. In addition, the KEGG genomic annotations are optionally utilized to extract the Enzyme Commission (EC) numbers for the putative enzymes in each organism. MetaCyc and other sources of functional annotation can also optionally be utilized. To further illustrate, one embodiment of the invention has data on 3,890 compounds, 2,917 transformations, and 100 organisms with annotations of putative gene functions.

C. IMPLEMENTATIONS

[0045] One embodiment of the invention has a modular and distributed architecture. There are two modes for interacting with this system: from the graphical user interface (GUI) through a web-browser, and from an interactive Common Lisp shell. The GUI is implemented as a Java client application that communicates with a Common Lisp server through TCP/IP using a custom Lisp protocol.

[0046] The server is implemented in Allegro Common Lisp and contains modules for data management, pathway inference by heuristic search, visualization and distributed computing. The purpose of the data management and pathway inference modules are described further above. The visualization module is responsible for producing a representation of the pathway suitable for rendering on the client. The distributed computing module is responsible for handling all client and server interaction, and for distributing the client requests across a Lisp Parallel Virtual Machine (McShan & Shah (2002) "Lisp-PVM: Parallel Virtual Machine in Lisp for Bioinformatics. Intelligent Systems for Molecular Biology" (Poster)).

1. WEB SITE EMBODIMENT

[0047] The methods of this invention can be implemented in a localized or distributed computing environment. For example, in one embodiment featuring a localized computing environment, a system of the invention comprises a computational device equipped with user input and output features. In a distributed environment, the methods can be implemented on a single computer, a computer with multiple processes or, alternatively, on multiple computers. The computers can be linked, e.g., through a shared bus, but more commonly, the computer(s) are nodes on a network. The network can be generalized or dedicated, at a local level or distributed over a wide geographic area. In certain embodiments, the computers are components of an intranet or an internet.

[0048] In such use, typically, a client (e.g., a scientist, practitioner, provider, or the like) executes a Web browser and is linked to a server computer executing a Web server. The Web browser is, for example, a program such as IBM's Web Explorer, Internet explorer, or the like. The Web server is typically, but not necessarily, a program such as IBM's HTTP Daemon or other WWW daemon (e.g., LINUX-based forms of the program). The client computer is bi-directionally coupled with the server computer over a line or via a wireless system. In turn, the server computer is bi-directionally coupled with a website (server hosting the website) providing access to software implementing the methods of this invention. A user of a client connected to the Intranet or Internet may cause the client to request resources that are part of the web site(s) hosting the application(s) providing an implementation of the methods of this invention. Server program(s) then process the request to return the specified resources (assuming they are currently available). A standard naming convention has been adopted, known as a Uniform Resource Locator ("URL"). This convention encompasses several types of location names, presently including subclasses such as Hypertext Transport Protocol ("http"), File Transport Protocol ("ftp"), gopher, and Wide Area Information Service ("WAIS"). When a resource is downloaded, it may include the URLs of additional resources. Thus, the user of the client can easily learn of the existence of new resources that he or she had not specifically requested.

[0049] Methods of implementing Intranet and/or Intranet embodiments of computational and/or data access processes are well known to those of skill in the art and are documented, e.g., in ACM Press, pp. 383-392; ISO-ANSI, Working Draft, "Information Technology-Database Language SQL", Jim Melton, Editor, International Organization for Standardization and American National Standards Institute, Jul. 1992; ISO Working Draft, "Database Language SQL-Part 2:Foundation (SQL/Foundation)", CD9075-2:199.chi.SQL, Sep. 11, 1997; and Cluer et al. (1992) A General Framework for the Optimization of Object-Oriented Queries, Proc SIGMOD International Conference on Management of Data, San Diego, California, Jun. 2-5, 1992, SIGMOD Record, vol. 21, Issue 2, Jun., 1992; Stonebraker, M., Editor. Other resources are available, e.g., from Microsoft, IBM, Sun and other software development companies.

Example Web Interface for Accessing Data Over a Network

[0050] Figure 5 A and B illustrate example interfaces for predicting biochemical pathways using a computer interface, possibly over a web page, according to specific embodiments of the present invention. Figure 5A illustrates the display of a Web page or other computer interface for requesting a biochemical pathway prediction. According to specific implementations and/or embodiments of the present invention, this example interface is sent from a server system to a client system when a user accessed the server system. This example Web page contains an input selection 501, allowing a user to specify input data. As will be understood in the art, each selection button can activate a set of cascading interface screens that allows a user to select from other available options or to browse for an input file. According to specific embodiments of the present invention, option selection 502 can also be provided, allowing a user to modify the user settable options discussed herein. A licensing information section 503 and user identification section 504 can also be included. One skilled in the art would appreciate that these various sections can be omitted or rearranged or adapted in various ways. The 504 section provides a conventional capability to enter account information or payment information or login information. (One skilled in the art would appreciate that a single Web page on the server system may contain all these sections but that various sections can be selectively included or excluded before sending the Web page to the client system.)

[0051] Figure 5B illustrates the display of an interface confirming a request. The confirming Web page can contain various information pertaining to the order and can optionally include a confirmation indication allowing a user to make a final confirmation to proceed with the order. For particular systems or analyses, this page may also include warnings regarding use of proprietary data or methods and can include additional license terms, such as any rights retained by the owner of the server system in either the data.

3. EMBODIMENT IN A PROGRAMMED INFORMATION APPLIANCE

[0052] Figure 6 is a block diagram showing a representative example logic device in which various aspects of the present invention may be embodied. As will be understood to practitioners in the art from the teachings provided herein, the invention

can be implemented in hardware and/or software. In some embodiments of the invention, different aspects of the invention can be implemented in either client-side logic or server-side logic. As will be understood in the art, the invention or components thereof may be embodied in a fixed media program component containing logic instructions and/or data that when loaded into an appropriately configured computing device cause that device to perform according to the invention. As will be understood in the art, a fixed media containing logic instructions may be delivered to a viewer on a fixed media for physically loading into a viewer's computer or a fixed media containing logic instructions may reside on a remote server that a viewer accesses through a communication medium in order to download a program component.

[0053] Figure 6 shows an information appliance (or digital device) 600 that may be understood as a logical apparatus that can read instructions from media 617 and/or network port 619, which can optionally be connected to server 620 having fixed media 622. Apparatus 600 can thereafter use those instructions to direct server or client logic, as understood in the art, to embody aspects of the invention. One type of logical apparatus that may embody the invention is a computer system as illustrated in 600, containing CPU 607, optional input devices 609 and 611, disk drives 615 and optional monitor 605. Fixed media 617, or fixed media 622 over port 619, may be used to program such a system and may represent a disk-type optical or magnetic media, magnetic tape, solid state dynamic or static memory, etc. In specific embodiments, the invention may be embodied in whole or in part as software recorded on this fixed media. Communication port 619 may also be used to initially receive instructions that are used to program such a system and may represent any type of communication connection.

[0054] The invention also may be embodied in whole or in part within the circuitry of an application specific integrated circuit (ASIC) or a programmable logic device (PLD). In such a case, the invention may be embodied in a computer understandable descriptor language, which may be used to create an ASIC, or PLD that operates as herein described.

4. INTEGRATED SYSTEMS

[0055] Integrated systems, e.g., for the methods described herein, as well as for the compilation, storage and access of databases, typically include a digital

computer with software including an instruction set as described herein, and, optionally, one or more of control software, analysis software, other data interpretation software, an input device (e.g., a computer keyboard) to enter data to the digital computer, to control analysis operations, etc.

[0056] Readily available computational hardware resources using standard operating systems can be employed and modified according to the teachings provided herein, e.g., a PC (Intel x86 or Pentium chip- compatible DOS™, OS2™, WINDOWS™, WINDOWS NT™, WINDOWS 95™, WINDOWS 98™, WINDOWS 2000™, WINDOWS XP™, LINUX, or even Macintosh, Sun or PCs will suffice) for use in the integrated systems of the invention. Current art in software technology is adequate to allow implementation of the methods taught herein on a computer system. Thus, in specific embodiments, the present invention can comprise a set of logic instructions (either software, or hardware encoded instructions) for performing one or more of the methods as taught herein. For example, software for providing the biochemical pathway predictions can be constructed by one of skill using a standard programming language such as Common Lisp, Visual Basic, Fortran, Basic, Java, or the like. Such software can also be constructed utilizing a variety of statistical programming languages, toolkits, or libraries.

[0057] Various programming methods and algorithms, including genetic algorithms and neural networks, can be used to perform aspects of the data collection, correlation, and storage functions, as well as other desirable functions, as described herein. In addition, digital or analog systems such as digital or analog computer systems can control a variety of other functions such as the display and/or control of input and output files. Software for performing the methods of the invention, such as programmed embodiments of the methods described above, are also included in the computer systems of the invention. Alternatively, programming elements for performing such methods as principle component analysis (PCA) or least squares analysis can also be included in the digital system to identify relationships between data. Exemplary software for such methods is provided by Partek, Inc. (St. Peter, MO); on the world wide web at partek.com.

Example System Embodiment

[0058] Figure 7 is a block diagram illustrating components that can be included in an integrated system according to specific embodiments of the present invention. This particular example embodiment optionally supports providing biochemical pathway predictions over a network. The server system 710 includes a server engine 711, various interface pages 713, data storage 714 for storing instructions, data storage 715 for storing, e.g., state data, state-transition data, etc., and data storage 716 for storing data generated by the computer system 710. According to specific embodiments of the invention, the server system further includes or is in communication with a processor 740 that further comprises one or more logic modules for performing one or more methods as described herein.

[0059] Optionally, one or more client systems may also comprise any combination of hardware and/or software that can interact with the server system. These systems may include digital workstation or computer systems (an example of which is shown as 720a) including a logic interface module (such as 721a) and/or various other systems or products through which data and requests can be communicated to a server system. These systems may also include laboratory-workstation-based systems (an example of which is shown as 720b) including a logic interface module (such as 721b) or various other systems or products through which data and requests can be communicated to a server system.

5. OTHER EMBODIMENTS

[0060] The invention has now been described with reference to specific embodiments. Other embodiments will be apparent to those of skill in the art. In particular, a viewer digital information appliance has generally been illustrated as a personal computer. However, the digital computing device is meant to be any information appliance for interacting with a remote data application, and could include such devices as a digitally enabled television, cell phone, personal digital assistant, etc.

[0061] Although the present invention has been described in terms of various specific embodiments, it is not intended that the invention be limited to these embodiments. Modification within the spirit of the invention will be apparent to those skilled in the art. In addition, various different actions can be used to effect the methods

described herein. For example, a voice command may be spoken by the purchaser, a key may be depressed by the purchaser, a button on a client-side scientific device may be depressed by the user, or selection using any pointing device may be effected by the user.

[0062] It is understood that the examples and embodiments described herein are for illustrative purposes and that various modifications or changes in light thereof will be suggested by the teachings herein to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the claims.

D. EXAMPLES

[0063] This example refers to a biochemical state-space that was built using data from known enzyme-catalyzed transformations in Ligand (Goto et al. (2002) "LIGAND: database of chemical compounds and reactions in biological pathways," Nucleic Acids Res 30(1):402-404), including, 2,917 unique transformations between 3,890 different compounds. To predict biochemical pathways this state-space was explored using an informed search algorithm that implements a chemically motivated heuristic to guide the search. Since the algorithm does not depend on predefined pathways, it can efficiently identify plausible routes using known biochemical transformations.

[0064] More specifically, this example provides the results of searching sample biochemical pathways using a computer implemented embodiment of a method according to the present invention. First, the efficiency of uninformed versus heuristic search in biochemical state-space is compared. Then, a brief overview of using the web-based system and a description of the pathway visualization is provided.

[0065] The efficiency of three different search algorithms in the state-space was compared including, breadth first search, depth first search and heuristic search, which were described further above. The results for four sample searches are summarized in Table 1. The predicted pathways include examples of biodegradation, biosynthesis, and biochemical engineering. For each of the pathway searches, the number of states explored (M), the length of the pathway (L), the cost (F), the effective branching factor (b*), and the computation time are provided. The computation time was measured by the Common Lisp function, called "time", which reports on the CPU usage of any computation. The timing was carried out by conducting the searches in the

interactive Common Lisp shell. The timing reported in Table 1 was measured using Allegro Common Lisp running in the RedHat Linux 7.3 operating system on a Sony PCG-CIMW laptop, containing a Transmeta Crusoe TM5800 Processor and 384 MB of main memory.

[0066] The pathway queries in Table 1 cover three kinds of biochemical themes including, biodegradation (example (a)), biosynthesis (examples (b) and (d)), and engineering (example (c)). The efficiency of a web-browser, and from an interactive Common Lisp the heuristic search algorithm is described in example (a). The exploration of the pathway from α-D-glucose to pyruvate yielded a large number of possible solutions using blind search strategies. Table 1 illustrates the efficiency of heuristic search in state-space over blind search strategies. Although breadth-first search finds the shortest path, it is the least efficient because it explores the largest number of states. Depth-first search is more efficient since lesser nodes are explored, but the drawback is that it produces a much longer path. The last column on the right shows the path found by heuristic search, which is the most efficient in the *F*-cost but not always the shortest. The quantitative performance measures for each of the search methods are also summarized in Table 1. Other analyses of other pathway searches using these three methods have also been performed (data not shown) and the A* was the most efficient in *F* cost and in the number of states explored. Though breadth-first gives the shortest path length, L, A* is the most efficient in exploring the state-space and in optimizing the *F* cost. The effective branching factor *b** is another useful metric for comparing the searches. The breadth-first search had the highest branching factor (2.27) because it explored all immediate successors first; the depth-first search had the lowest *b** (1.28) closely followed by A* (1.38). Though the branching factor of depth-first search is the lowest, it produces very long pathways, which do not seem to be very plausible. The time required for each search is roughly proportional to M.

[0067] To further illustrate the invention, Figure 8 schematically shows the visualization of a linear pathway. In particular, from top to bottom the figure shows: the profile of the F-cost along the steps in the pathway; the successors of each state are shown as points; the chemical structure of the main compounds at each step; helpful statistics about each step in the pathway; and the EC number of the enzymes involved in

catalyzing the transformations. The lower part of the visualization shows the compound names, their state descriptors, and the name of the enzyme.

[0068] While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. For example, all the techniques and apparatus described above may be used in various combinations. All publications, patents, patent applications, or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent application, or other document were individually indicated to be incorporated by reference for all purposes.

Example (a)
From: α-D-Glucose
To : pyruvate

Alg	M	L	F	b*	t(s)
BFS	209	5	29	2.65	8.86
DFS	26	22	491	1.01	1.64
A*	27	6	29	1.44	1.83

Example (b)
From: citrate
To : L-tyrosine

Alg	M	L	F	b*	t(s)
BFS	126	3	37	4.63	6.76
DFS	4718	123	2013	2.00	69.14
A*	20	6	13	1.34	1.98

Example (c)
From: α-D-glucose
To : 1,3-propanediol

Alg	M	L	F	b*	t(s)
BFS	652	7	81	2.33	19.42
DFS	19	17	185	1.01	1.83
A*	112	7	31	1.74	6.90

Example (d)
From: citrate
To : L-histidine

Alg	M	L	F	b*	t(s)
BFS	653	6	45	2.73	16.97
DFS	-	-	-	-	$> 10^4$
A*	72	7	15	1.61	4.20

Table 1: The tables provided above show the performance of the different search algorithms in state-space for four sample queries. Each table summarizes the results for a query from a starting compound to a goal compound. In each table, the columns contain the statistics for the computed pathways and the rows compare the results for the different search algorithms. In each table, from left to right the columns contain the algorithm (Alg); the total number of states explored in the search (M); the length of the path (L); the path cost *(F)*, which is calculated using equation 8; the effective branching factor (b*), which is computed using equation 9; and the time required for the computation (t). For each table, the second row shows the results for breadth first search (BFS); the third row for depth first search (DFS); and the fourth row for A* search (A*).

WHAT IS CLAIMED IS:

1. A method of predicting a biochemical pathway, the method comprising:

providing a population of compounds, which population comprises one or more input compounds and one or more output compounds;

defining at least one state-space that comprises the population of compounds; and,

identifying one or more candidate biochemical pathways between at least one of the input compounds and at least one of the output compounds using at least one informed search technique to search the state-space, thereby predicting the biochemical pathway.

2. A computer program product comprising a computer readable medium having one or more logic instructions for:

receiving data that defines at least one state-space comprising a population of compounds, which population comprises one or more input compounds and one or more output compounds; and,

identifying one or more candidate biochemical pathways between at least one of the input compounds and at least one of the output compounds using at least one informed search technique to search the state-space.

3. A system for predicting a biochemical pathway, comprising at least one computer having system software comprising one or more logic instructions for:

receiving data that defines at least one state-space comprising a population of compounds, which population comprises one or more input compounds and one or more output compounds; and,

identifying one or more candidate biochemical pathways between at least one of the input compounds and at least one of the output compounds using at least one informed search technique to search the state-space.

METHODS, SYSTEMS, AND SOFTWARE FOR PREDICTING BIOCHEMICAL PATHWAYS

ABSTRACT OF THE DISCLOSURE

[0069] The present invention provides methods for predicting biochemical pathways. Related systems and software are also provided.

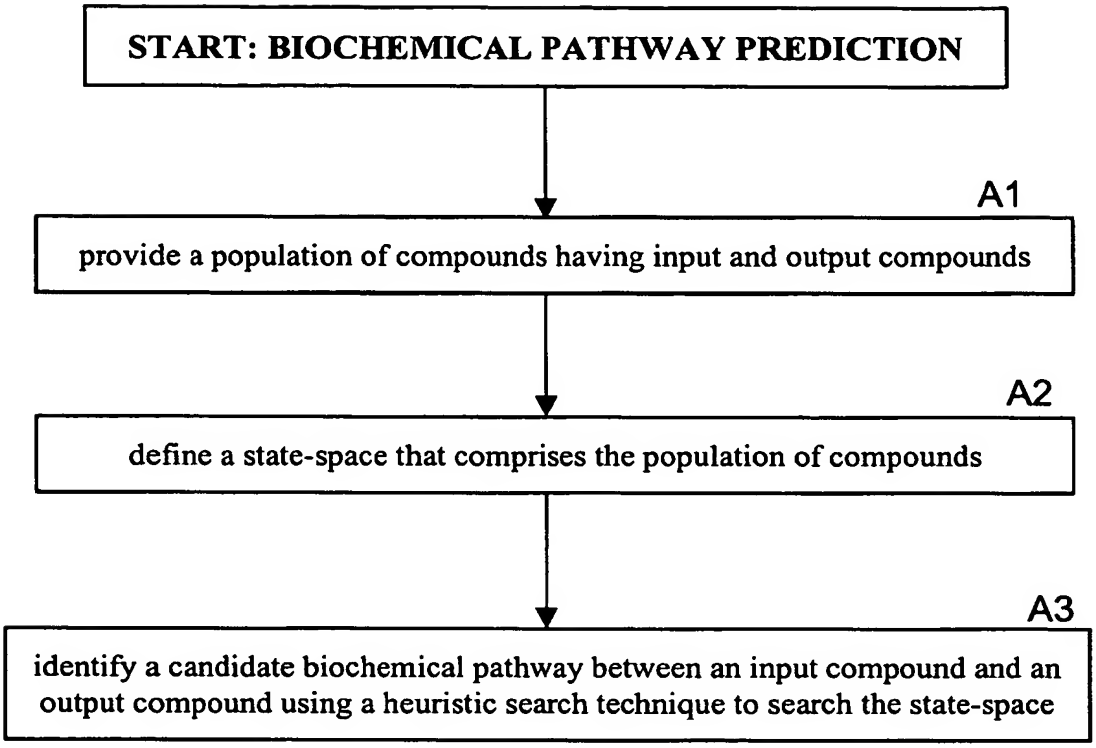

START: BIOCHEMICAL PATHWAY PREDICTION
A1
provide a population of compounds having input and output compounds
A2
define a state-space that comprises the population of compounds
A3
identify a candidate biochemical pathway between an input compound and an output compound using a heuristic search technique to search the state-space

Fig. 1

```
(AG AS AU BI BR C CA CD CL CO CU F FE GA H H+ HG I K LI MG MN MO N NA
 NI O P PB PT R S SB SE SI SN TC TE W X ZN BR-R C#C C#N C#O C-*
 C-AS C-BR C-C C-CL C-CO C-F C-FE C-H C-HG C-I C-MO C-N C-O C-P C-R C-S
 C-SE C-SI C-SN C-TC C-TE C-X C=C C=N C=O C=R C=S CL-CA CL-FE CL-HG
 CL-MG CL-PT CL-R H-CL H-N H-O HG-R I-I I-R K-I N#N N-* N-CO N-FE
 N-MG N-N N-NI N-O N-P N-PT N-R N-S N-SN N-X N-ZN N=N N=O N=P N=S O-*
 O-AS O-BI O-CA O-CL O-CO O-FE O-HG O-K O-MG O-NA O-O O-P O-R O-S O-SB
 O-SE O-SI O-SN O=AS O=CL O=O O=P O=S O=SE P-F P-S P=AU P=S P=SE S-AS
 S-AU S-F S-FE S-HG S-MO S-R S-S SE-SE)
```


Fig. 2

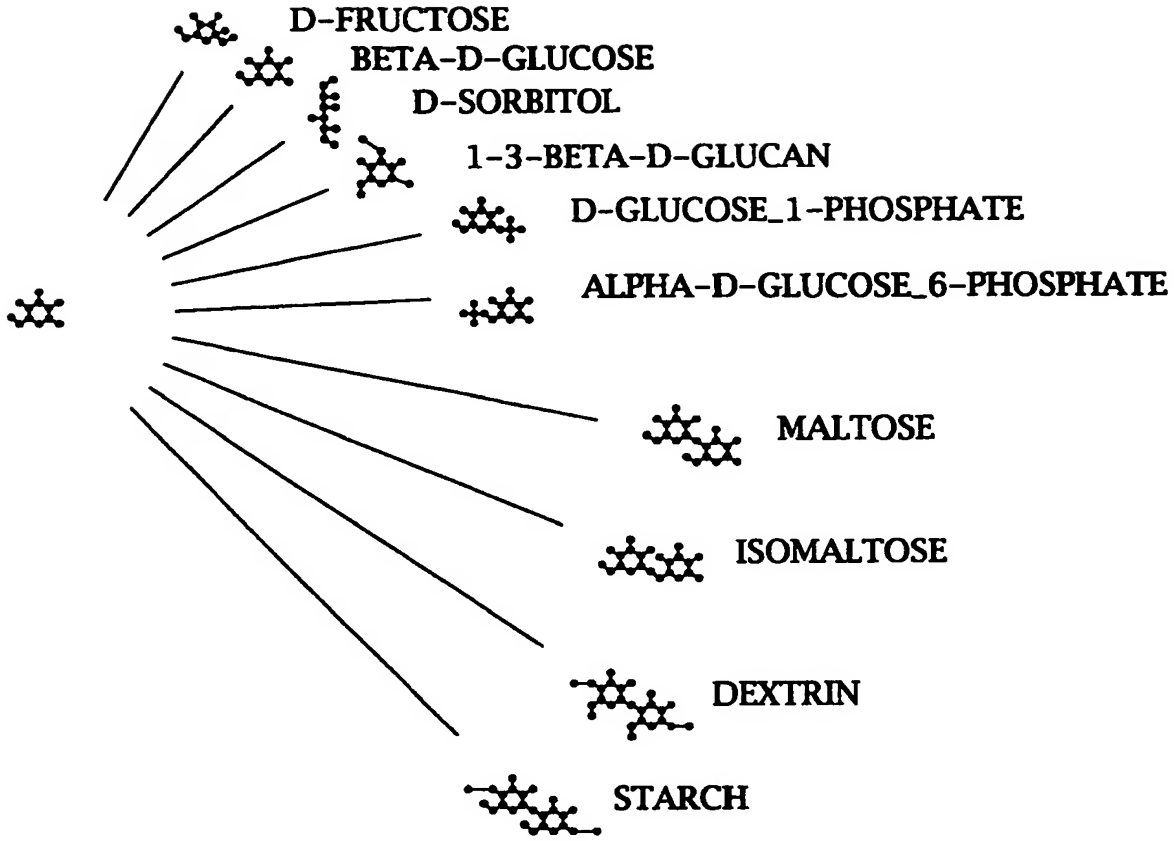

D-FRUCTOSE
BETA-D-GLUCOSE
D-SORBITOL
1-3-BETA-D-GLUCAN
D-GLUCOSE_1-PHOSPHATE
ALPHA-D-GLUCOSE_6-PHOSPHATE
MALTOSE
ISOMALTOSE
DEXTRIN
STARCH

Fig. 3

```
input  : x^0, x^L, F
output : P^{0,L}
begin
    X ← (x^0), P^{0,L} ← ()
    while X ≠ () do
        x ← argmax(F(x^i); x^i ∈ X)
        T ← successors(x)
        for x^m ∈ T do
            if x^m = x^L then
                P^{0,L} ← path(x^m)
                return P^{0,L}
            if x^m ∉ X then
                push (x^m, X)
                point(x^m, x)
            else
                if F(x^m) < F(x^m)|_old then
                    point(x^m, x)
end
```

Fig. 4

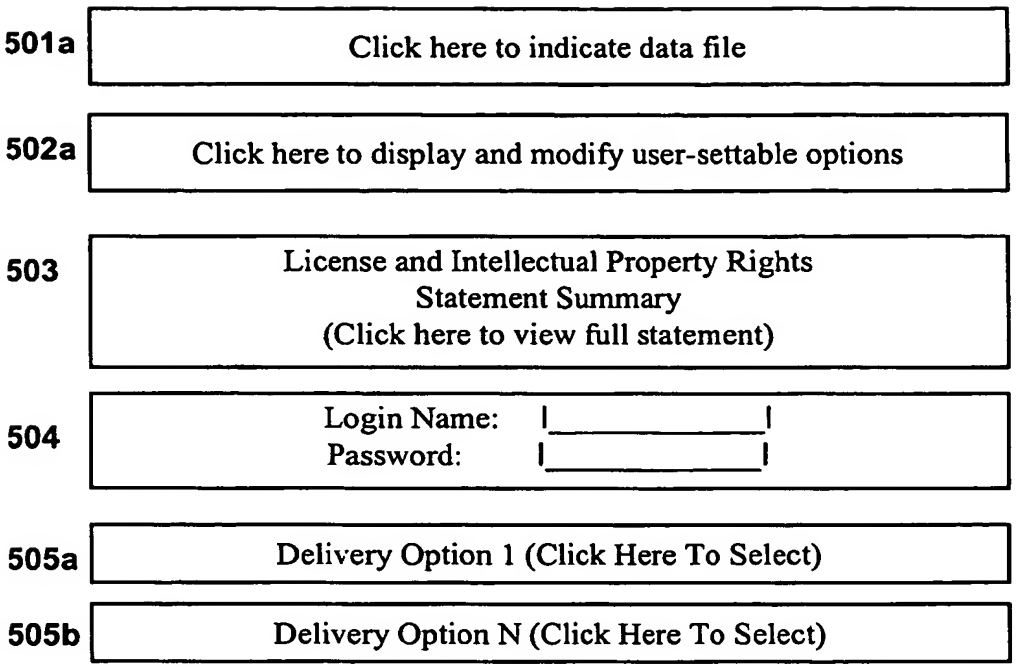

501a
Click here to indicate data file
502a
Click here to display and modify user-settable options
503
License and Intellectual Property Rights
Statement Summary
(Click here to view full statement)
504
Login Name: |_____|
Password: |_____|
505a
Delivery Option 1 (Click Here To Select)
505b
Delivery Option N (Click Here To Select)

Fig. 5A

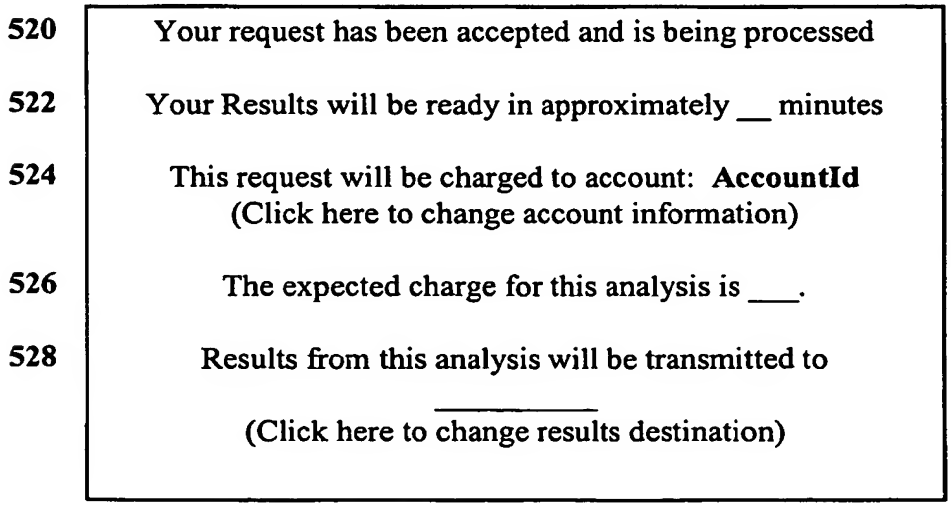

520
Your request has been accepted and is being processed
522
Your Results will be ready in approximately __ minutes
524
This request will be charged to account: **AccountId**
(Click here to change account information)
526
The expected charge for this analysis is ___.
528
Results from this analysis will be transmitted to
_____
(Click here to change results destination)

Fig. 5B

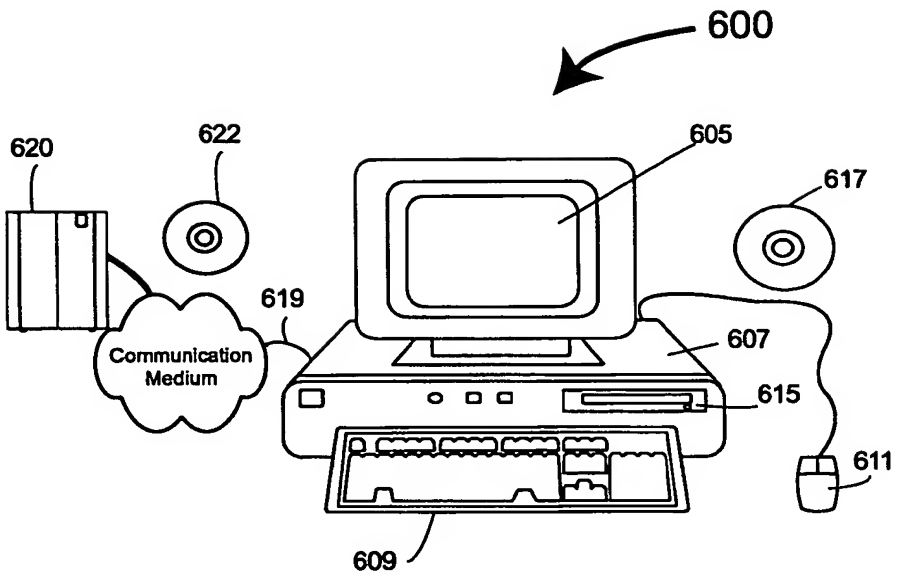


Fig. 6

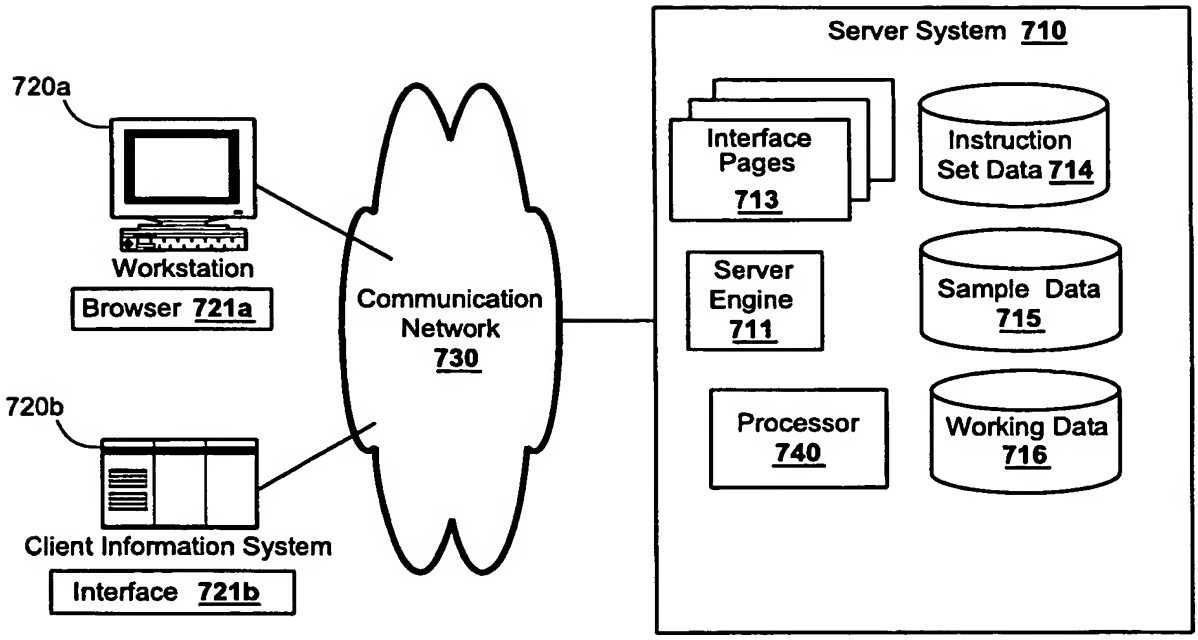
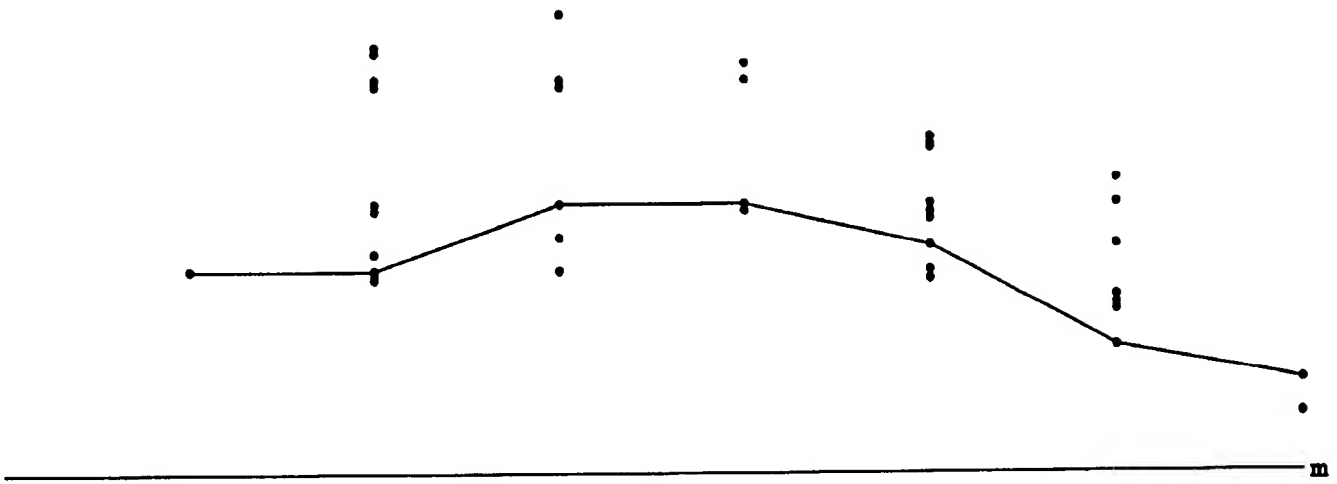

Server System 710
720a
Workstation
Browser 721a
Communication Network 730
720b
Client Information System
Interface 721b
Interface Pages 713
Instruction Set Data 714
Server Engine 711
Sample Data 715
Processor 740
Working Data 716

Fig. 7



m	0	1	2	3	4	5	6
Features	24	24	32	32	27	15	11
Successors	10	7	5	11	8	3	46
G(0,m)	0	0	8	8	13	25	29
H(m,L)	13	13	21	21	16	4	0
F(0,m,L)	13	13	29	29	29	29	29
EC		5.1.3.3	2.7.1.63	5.3.1.9	2.2.1.1	4.1.2.9	1.2.3.3

0 ALPHA-D-GLUCOSE	((C 6) (O 6) (C-C 5) (C-O 7))	ALDOSE_1-EPIMERASE
1 BETA-D-GLUCOSE	((C 6) (O 6) (C-C 5) (C-O 7))	POLYPHOSPHATE-GLUCOSE_PHOSPHOTRANSFERASE
2 BETA-D-GLUCOSE_6-PHOSPHATE	((C 6) (O 9) (P 1) (C-C 5) (C-O 7) (O-P 3) (O=P 1))	GLUCOSE-6-PHOSPHATE_ISOMERASE
3 BETA-D-FRUCTOSE_6-PHOSPHATE	((C 6) (O 9) (P 1) (C-C 5) (C-O 7) (O-P 3) (O=P 1))	TRANSKETOLASE
4 D-XYLULOSE_5-PHOSPHATE	((C 5) (O 8) (P 1) (C-C 4) (C-O 4) (C=O 1) (O-P 3) (O=P 1))	PHOSPHOKETOLASE
5 ACETYL_PHOSPHATE	((C 2) (O 5) (P 1) (C-C 1) (C-O 1) (C=O 1) (O-P 3) (O=P 1))	PYRUVATE_OXIDASE
6 PYRUVATE	((C 3) (O 3) (C-C 2) (C-O 1) (C=O 2))	

Fig. 8

Application Information

Application Type::	Provisional
Subject Matter::	
Suggested Classification::	
Suggested Group Art Unit ::	
CD-ROM or CD-R?::	
Number of CD disks::	
Number of copies of CDs::	
Sequence submission::	
Computer Readable Form (CRF)?::	
Number of copies of CRF::	
Title Line One::	METHODS, SYSTEMS, AND SOFTWARE
Title Line Two::	FOR PREDICTING BIOCHEMICAL
Title Line Three::	PATHWAYS
Attorney Docket Number::	61-000800US
Request for Early Publication?::	
Request for Non—Publication?::	
Suggested Drawing Figure::	
Total Drawing Sheets::	08
Small Entity::	Small
Petition included?::	
Petition Type::	
Secrecy Order in Parent Appl.?::	

Applicant Information

Applicant Authority Type::	Inventor
Primary Citizenship Country::	United States
Status::	
Given Name::	Imran
Middle Name::	
Family Name::	Shah
Name Suffix::	
City of Residence::	Denver
State or Province of Residence::	Colorado
Country of Residence::	United States
Street of mailing address::	1230 Ivanhoe Street
City of mailing address::	Denver
State of mailing address::	Colorado
Country of mailing address::	United States
Postal or Zip Code of mailing address::	80220

Applicant Authority Type::	Inventor
Primary Citizenship Country::	United States
Status::	
Given Name::	Daniel
Middle Name::	Charles
Family Name::	McShan
Name Suffix::	
City of Residence::	Littleton
State or Province of Residence::	Colorado
Country of Residence::	United States
Street of mailing address::	2901-D West Long Drive
City of mailing address::	Littleton
State of mailing address::	Colorado
Country of mailing address::	United States
Postal or Zip Code of mailing address::	80120

Correspondence Information

Correspondence Customer Number::	22798
Name::	Quine Intellectual Property Law Group
Street of mailing address::	P.O. Box 458
City of mailing address::	Alameda
State or Providence of mailing address::	California
Country of mailing address::	U.S.A.
Postal or Zip Code of mailing address::	94501
Phone number::	(510) 337-7871
Fax number::	(510) 337-7877
E-Mail address::	jaquine@quinelaw.com

Representative Information

Representative Customer Number::	22798

Assignee Information

Assignee Name:: The Regents of the University of Colorado

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US04/023566

International filing date: 21 July 2004 (21.07.2004)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/489,317
Filing date: 22 July 2003 (22.07.2003)

Date of receipt at the International Bureau: 18 April 2005 (18.04.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse